Sinhala Sentiment Analysis Using Comments Review

M.U.R. Mohamed, S.M.N.D. Bandara, A.G. Lakmal, and W.A.S.C. Perera

Department of ICT, Faculty of Technological Studies, University of Vavuniya, Sri Lanka

Abstract

Sinhala remains underrepresented in Natural Language Processing research, even though it is the primary language of more than 16 million people in Sri Lanka. To extract public opinion from the increasing amount of Sinhala user generated content on digital platforms, effective sentiment analysis tools are required. The study offers a comprehensive overview of sentiment analysis research on social media data and Sinhala news comments. The review discussed the available datasets, model optimization tactics, and algorithmic approaches of Sinhala sentiment analysis by analyzing thirteen research papers. The study examines a range of algorithms, including transformer based models such as SinBERT, XLM-R, and RoBERTa, ensemble methods, deep learning architectures like CNN, LSTM, BiLSTM, S-LSTM, and Capsule Networks, and traditional machine learning approaches such as Naïve Bayes, SVM, and Logistic Regression. The results show that although multilingual transformers such as XLM-R and SinBERT have shown remarkable effectiveness, several non-transformer models including BiLSTM and Logistic Regression, have also performed well on particular tasks. However, significant challenges remain, including the handling of code-mixed language, the morphological complexity of Sinhala, and the scarcity of annotated data. This review highlights these important research gaps and suggests potential directions for Sinhala sentiment analysis.

Keywords: Natural Language Processing, News Comments, Sentiment Analysis, Sinhala

Department of ICT, Faculty of Technological Studies, University of Vavuniya 2025. All rights reserved